

APPLICATION OF A COMPUTERIZED GENERAL PURPOSE
INFORMATION MANAGEMENT SYSTEM (SELGEM) TO
A NATURAL HISTORY RESEARCH DATA BANK
(COLEOPTERA: CARABIDAE)

LA VERNE J. MAGARIAN ERWIN

c/o Department of Entomology, National Museum of Natural History,
Smithsonian Institution, Washington DC 20560

ABSTRACT

Natural history data can be computerized to aid the systematics researcher by rearranging data for synthesis and analysis and in preparation of manuscripts. This paper describes the processes of data capture in the field and the work flow procedures to get this data onto computer readable magnetic tape and into a data bank. Field note standardized format is exemplified and discussed. Sixty-two data categories (data elements) are defined and discussed. This system was designed for use in a Central American Ground Beetle Faunal Study (Coleoptera: Carabidae), but can easily be adapted to other organisms. SELGEM, a general purpose information management system, is used. The data resources for the Faunal Study are field notes, specimen labels, specimens, and literature; the processes of literature data capture will be described elsewhere.

INTRODUCTION

Computer technology to aid systematists in accumulating, organizing, and evaluating the massive quantity of data with which they need to work has been available for several years. Data are from sources such as specimens, specimen labels, field notes, and literature. By using a computer, data only need be accumulated once if originally entered correctly for output. Once accumulated, and if appropriate computer programs are available, the computer can permute these data into lists, maps, indices, camera ready copy, graphs, etc. Printed copies of extracted and/or organized data (in any combination) can be used to assist systematists in the analytical and synthetic portions of their work, as a working or final report, for reference, or can be sent to colleagues.

Computer technology can aid the systematist, but the systematist must recognize the restrictions which computer technology entails; computer technology is not capable of organizing unstructured, nonconsistent raw data. *"If there is any one lesson to be learned from a data-banking experience, it is that a high premium should be placed on entering the data correctly in the first place, down to the finest detail in formatting, spacing, spelling, abbreviating and punctuating"* (Shetler 1974). Computer technology does not eliminate manual hours, but needs additional manual hours for specimen data preparation, as well as time for editing computer processed data. (See Shetler 1974 for a "devil's advocate" approach to computer technology as applied to biological data banking). If systematists have available the experience from other studies, they will be able to begin each subsequent study at a higher, more refined level; or if the previous study's

level of refinement is applied, the data can be more quickly built into a meaningful data base.

For several years, the NMNH (Smithsonian Institution) has been involved in the development and use of procedures for computerization of data contained in systematic collections (Squires 1966, 1970; Manning 1969; Gautier 1974). The computer system developed at the Smithsonian to incorporate computer technology with museum needs is SELGEM, a general purpose information management system¹ (Creighton and Crockett 1971; Creighton, Packard, and Linn 1972; Creighton and Packard 1974; Chenhall 1975; Creighton and Robinson 1975; R. Creighton, P. Packard, and J. J. Crockett, unpublished manuscript).² Some advantages of SELGEM are that data records can be entered in fixed or variable length formats; SELGEM categories (see glossary) of data can be created simply by adding a unique 3 digit number to the file; data can be entered by paper tape (any code structure), port-a-punch cards, punched cards, or typed copy (Optical Character Recognition [OCR], see glossary); characters within a line, and lines within a SELGEM category can be added, modified, or deleted at any time; records can be added or deleted at any time; repetitious additions, modifications, and deletions need be entered only once; file cards, labels, reports (in formats specified by the user), and summarized reports with or without totals can be printed; indices can be made by word or phrase; and data files can be interfaced with map programs, photo-composition equipment, and statistical programs (see overview of SELGEM capabilities by D. Bridge, unpublished manuscript available through the NMNH ADP Program¹).

This paper describes the NMNH Central American Ground Beetle (Carabidae s. l.) Faunal Study (hereafter referred to as "the Study"), which uses the SELGEM system. The Study (whose primary outcome will be a series of faunal manuals and subsequent analyses) has successfully applied SELGEM to various research requirements. After 3½ years of refining the Study, compatibility has been achieved in all aspects from field note to data entry; this paper describes the capturing of data *from* field notes designed especially for the Study's output requirements *to* the entering of data for computer processing. It is hoped that future studies can benefit from the Study's present level of refinement. If data bases are built using the same computer system, definitions, and formats, more opportunities for uniform exchange of data about world faunas will exist.

The scientific work for this study is being done by Terry L. Erwin, Smithsonian Institution. The design of the work flow from field note to formats for encoding data, and the actual building of the data base, are handled by me. The Smithsonian's NMNH ADP Program and Office of Computer Services¹ do the computer processing and serve as SELGEM consultants.

In 1972 the Study shifted from conventional procedures of handling data (recording notes and literature citations on 5 by 7 inch cards, recording with pen or pencil lengthy field notes, and manually sorting and resorting specimens and cards for various needs) to using the computer as an aid in storage and retrieval of the massive quantity of data. The original plan was simply to build 2 computer files, one of specimen data and one of literature data, merge them, and then use the data in various combinations to aid in the faunal Study. The original plan remains, but the impact of the computer approach on the entire Study has been far greater than originally expected.

Particularly influenced is the structure and content of field notes, since

in this early phase of research key data are first gathered and recorded. Formerly, field notes were written in an unstructured, unformatted, narrative style. Locality data, elevation, and general microhabitat were always entered into the notes; but ecological data such as community, macrohabitat, time of activity, etc., were randomly included, word synonymies were abundant, words and phrases lacked consistency in structure and definition (e.g. rot, rotting, rotten, decayed, decaying), and data were written in narrative paragraphs. These notes were not adaptable to the computer, since it works only with precise and consistent data. Also, retrievability of pertinent data from these notes was often dependent upon experience and memory recall, and many good field observations simply were forgotten for a lack of logical accumulation and organization. To use the computer, it was necessary to refine, format, and standardize field notes in order to maximize the efficiency, accuracy, and reliability of the data file.

Field notes are written on a printed form (Fig. 1), using a pocket-sized tape recorder for recording all required data on site. The 5½ by 8½ inch printed forms are kept in a 6½ by 9 inch loose leaf binder. A consecutive numbering system is used, providing each field note with a separate number for each specific microhabitat. One collecting vial is used for each microhabitat. In the evening, the specimens are transferred into storage vials of hardwood sawdust (Lindroth 1957) and field notes are transcribed for each microhabitat from the tape. Field identifications are rechecked and listed at the bottom and backside of each field note.

The printed form was designed to maximize efficiency and accuracy in recording notes. Specific locations are provided on the form for each kind of data item (see glossary) to be recorded. Where choices of items are few, key words are printed; the desired word(s) are encircled rather than written. The printed form enhances accuracy and eliminates the possibility of subsequent interpretation error, since the predetermined locations of data and the printed key words provide reminders of which data are to be recorded.

One or several forms might be filled out at each locality, depending upon the number of microhabitats investigated. On the first form such general information about the locality as country, province, etc. is recorded. Each subsequent form includes data on a microhabitat, and on the bottom and backside of the page a list of the taxa and number collected from the microhabitat. Typing data for computer entry is greatly facilitated by having a separate form for each microhabitat; this results in detailed field notes which contain most, if not all, of the data elements (see glossary) needed for later analysis.

Since the data elements which follow were specifically chosen and defined for the Central American Ground Beetle Faunal Study, there are not many gaps in data items computerized for specimens collected specifically for the Study; but there are many such gaps in the records of borrowed specimens. It has been possible in certain cases to locate the collector's field notes and find data not on the specimen labels. Also, certain data can be researched, such as latitude-longitude, province, or Holdridge Life Zone.

Appropriateness of the format and definitions of many of the data elements, especially those dealing with ecology, are still being tested. As the Study develops, many of these data elements will be expanded, dropped, or refined.

Codes and abbreviations have been used sparingly; only mnemonic

(800)—DATE—ΔΔ

SPECIMEN COUNT FOR FIELD NOTE—

(820)—FIELD NOTE NUMBER—

RUNNING COUNT FOR EXPEDITION—

(420)—COUNTRY—(circle one) BELIZE COSTA RICA EL SALVADOR GUATEMALA HONDURAS NICARAGUA PANAMA

(430)—DEPARTMENT/DISTRICT/PROVINCE—

(440)—GENERAL LOCALITY—

(450)—SPECIFIC LOCALITY—

(460)—LATITUDE-LONGITUDE COORDINATES—ΔΔN., ΔΔW.

(470)—ELEVATION—ΔTO ΔΔM

(810)—COLLECTOR(S)—

(500)—HOLDRIDGE LIFE ZONE—

(520)—COMMUNITY—(circle one) CAVE CUL-L FAL-F FORES GRASL INHAB PASTL SAVAN SCRUB SUBPA
(525)—COMMUNITY TYPE—
(526)—COMMUNITY TYPE NARRATIVE—
(560)—BORDERED BY—(circle one) CAVE:Δ CUL-L: FAL-F: FORES: GRASL: INHAB: PASTL: SAVAN:
SCRUB: SUBPA: RIPAR: SEEPG: LAKE:Δ MARSH: POND:Δ SWAMP: SEABE: ROADS:
(COMMUNITY TYPE)—
(561)—BORDERED BY—(circle one) CAVE:Δ CUL-L: FAL-F: FORES: GRASL: INHAB: PASTL: SAVAN:
SCRUB: SUBPA: RIPAR: SEEPG: LAKE:Δ MARSH: POND:Δ SWAMP: SEABE: ROADS:
(COMMUNITY TYPE)—

(520)—COMMUNITY—(circle one, record all of the following six categories) RIPAR SEEPG
(circle one, size and flow not applicable) LAKE MARSH POND SEABE SWAMP
(530)—SIZE—ΔTO ΔΔ M
(535)—FLOW—BABBLING CASCADING RAPID STANDING SLOW WATERFALL
(540)—SUBSTRATE—CLAY COBBLE GRAVEL HUMUS LOAM SAND SILT STONES
(541)—SUBSTRATE NARRATIVE—
(545)—VEGETATION—
(560)—BORDERED BY—(circle one) CAVE:Δ CUL-L: FAL-F: FORES: GRASL: INHAB: PASTL: SAVAN:
SCRUB: SUBPA: RIPAR: SEEPG: LAKE:Δ MARSH: POND:Δ SWAMP: SEABE: ROADS:
(COMMUNITY TYPE)—

(520)—COMMUNITY—ROADS
(540)—SUBSTRATE—CLAY COBBLE GRAVEL HUMUS LOAM SAND SILT STONES
(541)—SUBSTRATE NARRATIVE—
(550)—STRUCTURE—DITCH: MOUND: FLAT:Δ narrative
(555)—ROADSIDE VEGETATION—P-ZONE: S-GROW: CLIMAX: narrative
(560)—BORDERED BY—(circle one) CAVE:Δ CUL-L: FAL-F: FORES: GRASL: INHAB: PASTL: SAVAN:
SCRUB: SUBPA: RIPAR: SEEPG: LAKE:Δ MARSH: POND:Δ SWAMP: SEABE: ROADS:
(COMMUNITY TYPE)—
(561)—BORDERED BY—(circle one) CAVE:Δ CUL-L: FAL-F: FORES: GRASL: INHAB: PASTL: SAVAN:
SCRUB: SUBPA: RIPAR: SEEPG: LAKE:Δ MARSH: POND:Δ SWAMP: SEABE: ROADS:
(COMMUNITY TYPE)—

(580)—MACROHABITAT—(circle one) CANOPY UNDERCANOPY SUPRA-EPIGEAN EPIGEAN HYPOGEAN

(600)—MICROHABITAT—(circle one) ANT COLUMN CARRION CLAY COBBLE DOMESTIC DUNG FLOWER FOLIAGE
FOLIAGE-EPIPHYTE FOLIAGE-FERNS FOLIAGE-GRASS FOLIAGE-HERBS FOLIAGE-MOSS
FOLIAGE-SEDGE FOLIAGE-SHRUBS FOLIAGE-TREES FOLIAGE-VINES FRUIT FUNGI GRAVEL
GUANO HUMUS LOAM LEAF LITTER MUD NEST ROOTS SAND SILT SOIL STONE
SPIDER WEB TREE FALL LEAF LITTER WOOD other
(605)—MICROHABITAT SPECIES NAME—
(610)—SECTION OF MICROHABITAT—
(615)—CONDITION OF MICROHABITAT—
(620)—STRATUM OF MICROHABITAT SECTION—
(625)—ACTIVITY OF SPECIMEN—
(630)—MICROHABITAT REMARKS—

(700)—COLLECTING METHOD—BEATING BERLESE FUNNEL RAKING SIFTING SPLASHING SWEEPING TREADING

(720)—ATTRACTANTS AND TRAPS—BAIT:Δ
TRAP:Δ
LIGHT:UV
LIGHT:UV, TRAP
LIGHT:UV, WHITE
LIGHT:WHITE

(730)—TIME OF ACTIVITY—Δ (Auroral Diurnal Vesperal Nocturnal)

(740)—EXPOSURE—Illum-OPEN~~N~~P-OPEN~~C~~CLOSED~~A~~ GRAD-FLAT~~A~~INTER~~A~~STEEP~~A~~ DIR-N E S W NE NW SE SW

(750)—TOPOGRAPHY—RAVINE~~A~~SEC-HEAD~~A~~MIND~~A~~MOUTH~~A~~ STR-SHALLOW~~A~~SLOPING~~A~~STEEP~~A~~ SIZE—ΔMA~~A~~ VEG—

(999)—MISCELLANEOUS REMARKS—

FIG. 1, Central American Field Note Form.

codes were used. First consideration is given to ease in usage and understandability; secondary consideration is given to savings in time. Since the Honeywell computer which the Smithsonian uses prints only upper case letters, the abbreviations are listed in upper case letters; periods are not used after abbreviations. Some abbreviations used in the Study are C=Centigrade; DET=Determined; E=East; F=Fahrenheit; HRS=Hours; HWY=Highway; JCT=Junction; KM=Kilometer(s); M=Meter(s); MI=Mile(s); MM=Millimeter(s); N=North; NR=Near; S=South; SP=Species; and W=West.

The cost of *computer processing* for entering specimen data and *making various queries and listings* using Procedure I³ averaged \$ 0.22 per specimen (average of 15 data lines per specimen) for 9,628 specimens. This figure includes experimentation costs, but it does not include such primary personnel costs as salary for personnel entering specimen data and arranging specimens for data entry. The cost of *computer processing* for entering specimen data *without queries and listings* using Procedures II and III³ averaged \$ 0.07 per specimen (average of 24 data lines per specimen) for 8,261 specimens. This figure does not include primary personnel costs. During the Study, the computer model used to process data was changed, computer programs changed constantly, charges for computer processing continually increased, and combinations of computer programs to process data varied. Therefore, these cost estimates are included only to give some idea of expenses necessary to undertake computerization of natural history data.

DATA CATEGORIES AND SELGEM CATEGORY NUMBERS USED IN THE CENTRAL AMERICAN CARABID NATURAL HISTORY STUDY^{4 5}

SPECIMEN NUMBER	010
DEPOSITORY	020
COLLECTION TRANSFER	030
SPECIMEN LOAN-RETURN	040
TRIBE	120
SUBTRIBE	130
GENUS	140
SUBGENUS	150
SPECIES	160
SUBSPECIES	170
IDENTIFIER BY TYPE COMPARISON	180
IDENTIFIER BY KEYS OR UNKNOWN METHOD OF COMPARISON	190
KIND OF TYPE	210
TYPE NUMBER	220
TYPE DESIGNATOR AND DATE	230
ORIGINAL TYPE NAME	240
STAGE OF METAMORPHOSIS	300
SPECIMEN MATURITY	310
SEX	320
SPECIMEN LENGTH	330
SPECIMEN WIDTH	340
WING LENGTH	350

ZOOGEOGRAPHIC REGION.....	400
COUNTRY	420
DEPARTMENT/DISTRICT/PROVINCE.....	430
GENERAL LOCALITY	440
SPECIFIC LOCALITY	450
LATITUDE-LONGITUDE	460
ELEVATION	470
HOLDRIDGE LIFE ZONE	500
COMMUNITY.....	520
COMMUNITY TYPE	525
COMMUNITY TYPE NARRATIVE	526
SIZE	530
FLOW.....	535
SUBSTRATE	540
SUBSTRATE NARRATIVE.....	541
VEGETATION	545
STRUCTURE	550
ROADSIDE VEGETATION	555
BORDERED BY	560
BORDERED BY	561
BORDERED BY	562
MACROHABITAT.....	580
MICROHABITAT	600
MICROHABITAT SPECIES NAME	605
SECTION OF MICROHABITAT	610
CONDITION OF MICROHABITAT	615
STRATUM OF MICROHABITAT SECTION.....	620
ACTIVITY OF SPECIMEN	625
MICROHABITAT REMARKS	630
COLLECTING METHOD.....	700
ATTRACTANTS AND TRAPS	720
TIME OF ACTIVITY.....	730
EXPOSURE	740
TOPOGRAPHY	750
DATE CAPTURED	800
COLLECTOR(S)	810
FIELD BOOK-NOTE NUMBERS AND HABITAT PHOTOGRAPH	820
COLLECTOR'S NUMBER.....	840
EXPEDITION.....	850
MISCELLANEOUS REMARKS	999

HOW DATA ARE PREPARED FOR COMPUTER PROCESSING

For this Study, source data (specimen, specimen labels, and field note data) are entered into paper tape or into Hollerith 80 column IBM punch cards. Most source data are entered via paper tape typewriters, but for some categories, designed so that data begin and end in specific character positions, the Hollerith punch card is used. The Hollerith punch card is also used to enter corrections to the file. Methods of data capture other than paper tape and Hollerith punch cards are also available.

The paper tape typewriter is composed of a (program) reader, a punch, and a standard electric typewriter. Punched paper tape programs are read

by the *program reader* which instructs the paper tape machine when to punch data, when to type data, when to stop for data entry, and so on. The *punch* punches selected data from the punched paper tape program and also punches the source data when they are typed.

When a reel of punched paper tape and/or a stack of punched cards (source data) are ready, Job Control Language (JCL) cards (also on punched cards) are made. The JCL cards instruct the computer to process the punched paper tape reel and/or punched cards with the appropriate SELGEM program, and also specify options which are available for each SELGEM program. After the JCL cards are made, the paper tape reel, Hollerith punched cards, and JCL cards are submitted to be processed.

PROCEDURES FOR PREPARING AND GATHERING SOURCE DATA

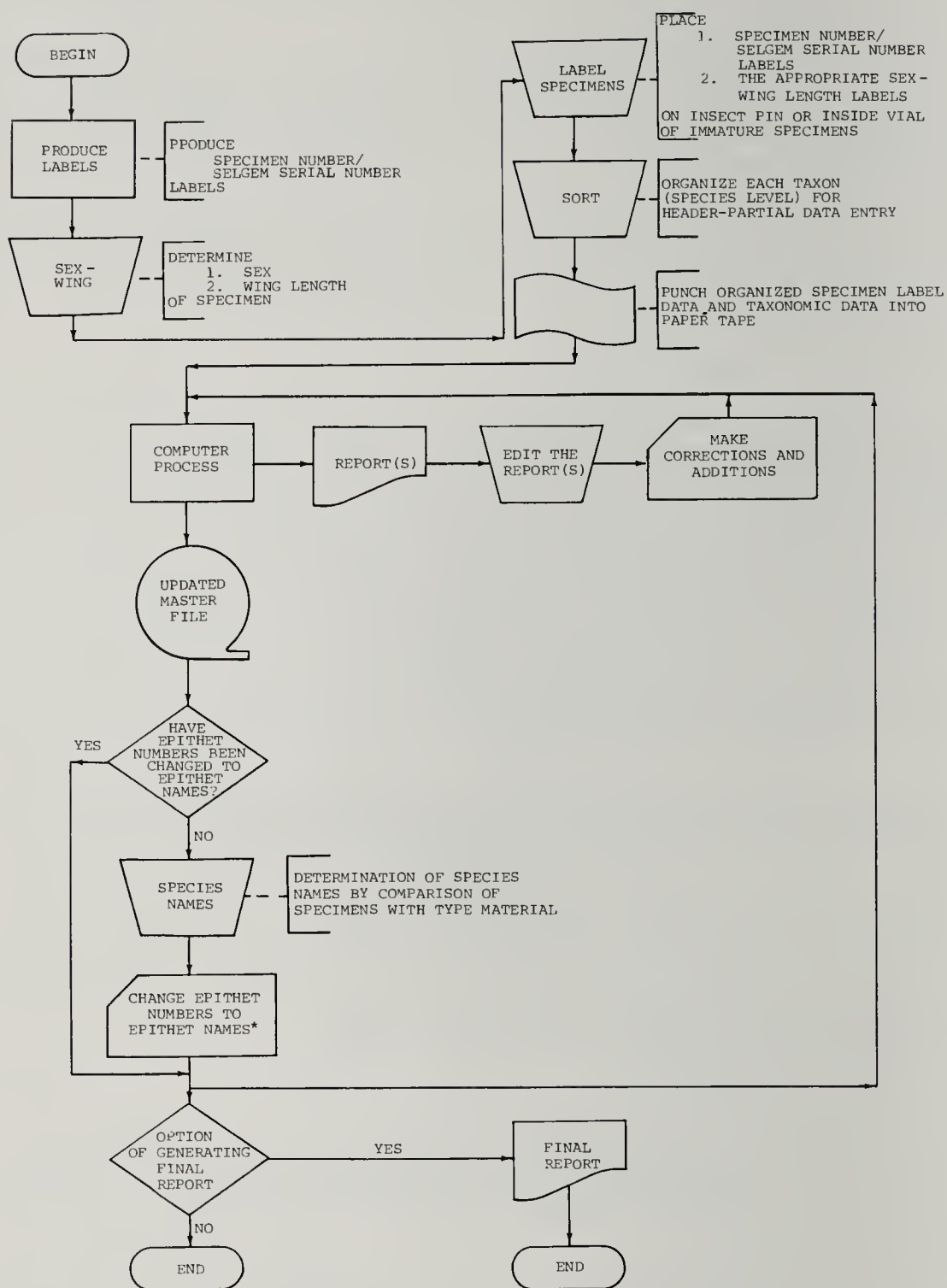
Data are taken directly from the specimens and specimen labels and are punched immediately into paper tape. Three different procedures for preparing and gathering source data for computer processing are used in this Study. Which of the 3 procedures is used depends upon how the collection is arranged prior to data entry. Non-detailed flowcharts graphically illustrate the procedures used (Fig. 2, 6, and 7). A legend at the bottom of Fig. 2 defines the symbols used in the flowcharts. The explanations for the following 3 procedures are intended to explain the general work flow, not to describe the procedures in minute detail.

PROCEDURE I (FIG. 2): THE USNM COLLECTION WHICH WAS IDENTIFIED AND ORGANIZED TO SPECIES PRIOR TO DATA ENTRY

- 1.0 Produce a bulk quantity of specimen number/SELGEM serial number labels (see Fig. 3 for examples of labels; see category 010 for explanation of number; see Footnote 6 for method to produce the specimen number/SELGEM serial number label).
- 2.0 Determine sex of specimen (category 320) and wing length of specimen (category 350).
- 3.0 Place specimen number/SELGEM serial number label and the appropriate sex-wing length label (Fig. 5) on insect pin or place specimen number/SELGEM serial number label inside alcohol vial of immature specimens.
- 4.0 Sort each taxon (species level⁷) so that constant data items from specimen labels are together, enabling use of header-partial data entry (see Glossary). Examples:
 - 4.1.1 Put all Costa Rican specimens together.
 - 4.1.2 Put all Costa Rican, Limon Province, males, winged specimens together.
- 5.0 Punch header-partial organized specimen label data and taxonomic data into paper tape.
- 6.0 Computer process data. Outputs of the processed data generally are:
 - 6.1 A new master file or an updated master file on magnetic tape.
 - 6.2 Working reports.
 - 6.2.1 The reports are edited.
 - 6.2.2 Corrections and additions are prepared (generally on punched cards) to update the master file (see glossary).
 - 6.2.3 The corrections are computer processed.

ERWIN: APPLICATION OF SELGEM

FLOWCHART OF PROCEDURE I:
THE USNM COLLECTION WHICH WAS
IDENTIFIED AND ORGANIZED TO SPECIES PRIOR TO DATA ENTRY



*THIS STEP IS NOT ALWAYS MADE THROUGH PUNCHED CARDS. SEE PROCEDURE I, STEP 7.1. FOR EXPLANATION

LEGEND:

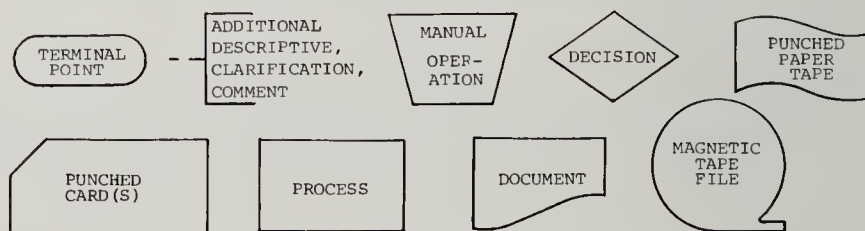


FIG. 2, Flowchart of Procedure I.

6.2.4 Steps 6.2.1 through 6.2.3 are repeated until all corrections and additions are satisfactorily completed.

7.0 Determination of species names by comparison of specimens with type material.

7.1 The epithet numbers are changed to epithet names either via punch cards or through various SELGEM programs. If the changes are made via punch cards, then steps 6.2.1 through 6.2.4 are repeated.

8.0 At this point the master file is basically finished. Working reports and/or final reports can be produced at any time. If latitude-longitude coordinates are included in the file, maps can be computer generated.

This was the first procedure used and also the slowest; as a result of using this procedure on 9,628 specimens, the other 2 methods were designed. If a collection is taxonomically arranged, the time spent to manually resort the specimens by locality for data entry ease, then resort the specimens back to the taxonomically arranged collection, in my experience, is not justified.

ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP
37238	37239	37240	37241	37242	37238	37239	37240	37241	37242
ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP
37260	37261	37262	37263	37264	37260	37261	37262	37263	37264
ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP
37282	37283	37284	37285	37286	37282	37283	37284	37285	37286

FIG. 3

ADP-USNM	ADP-CAS	ADP-MCZ	ADP-AMNH
19080820	19101101	19410908	19751224

FIG. 4

♀ winged	♀ short-winged	♀ wingless
♂ winged	♂ short-winged	♂ wingless

FIG. 5

FIG. 3-5, Labels: 3) ADP specimen number labels used presently by the Study; 4) Proposed specimen ADP number labels; 5) Sex/Wing specimen labels.

PROCEDURE II (FIG. 6): UNSORTED COLLECTIONS WHICH HAVE BEEN
BORROWED FROM PUBLIC (MUSEUM) AND/OR PRIVATE SOURCES

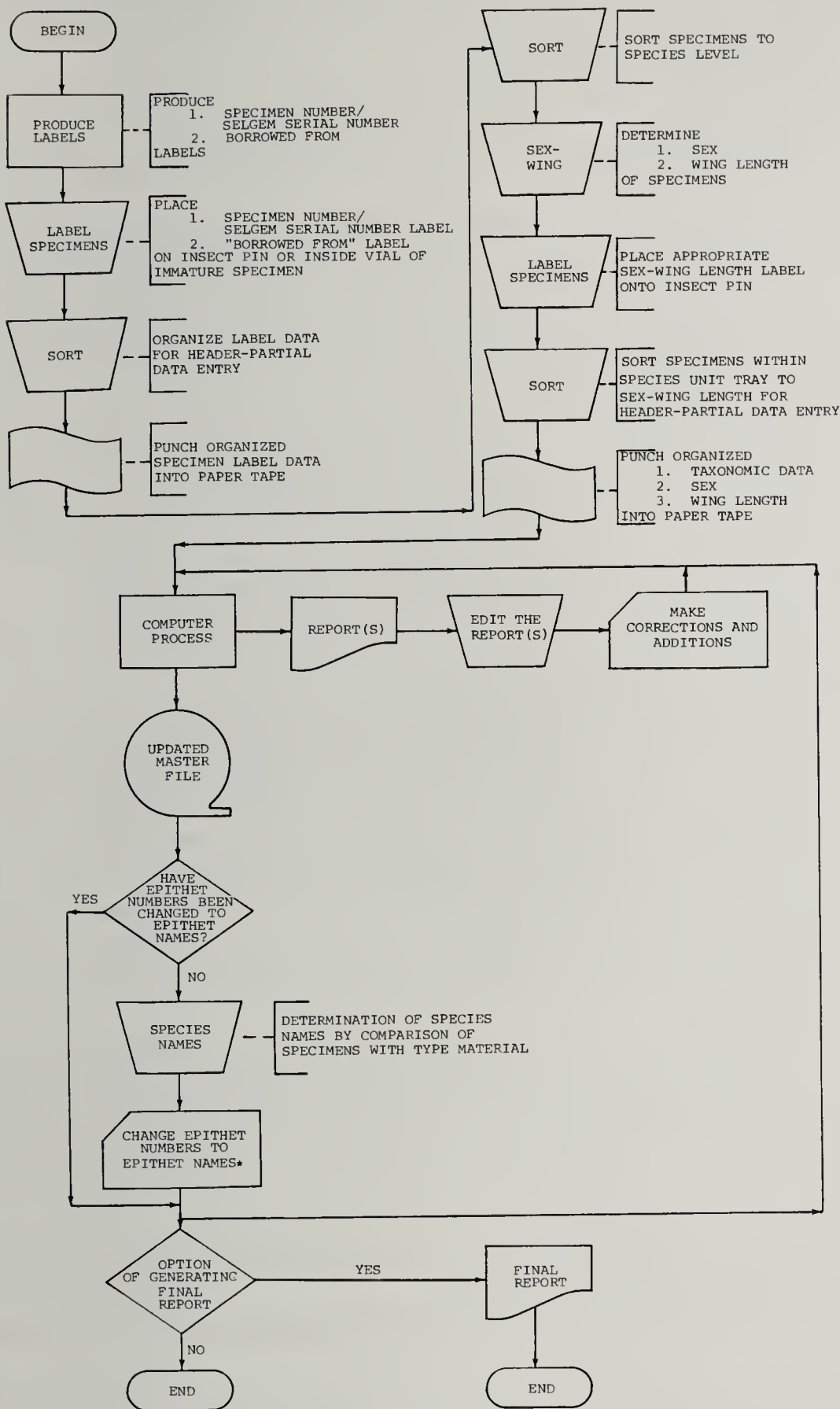
- 1.0 Produce labels
 - 1.1 Produce a bulk quantity of specimen number/SELGEM serial number labels (category 010) (step 1.0 in Procedure I).
 - 1.2 Produce "borrowed from" labels, if the specimens do not have adequate depository identification (category 020).
- 2.0 Label specimens
 - 2.1 Place specimen number/SELGEM serial number labels (category 010) on insect pins or inside alcohol vials of immature specimens.
 - 2.2 Place "borrowed from" labels on the insect pins which do not have adequate depository identification.
- 3.0 Sort specimens by label data so that constant data items are together, enabling header-partial data entry (steps 4.1.1, 4.1.2 in Procedure I). When punching the data into the paper tape, it is useful to slant the specimens in the unit tray (see glossary) as they are being sorted, so that the label data is seen at a glance.
- 4.0 Punch header-partial organized specimen label data into paper tape (categories such as 010, 020, 030, 210, 220, 230, etc.).
- 5.0 Sort specimens to species level.
 - 5.1 Label each unit tray with the appropriate Genus name and epithet number.⁷
- 6.0 Determine sex and wing length of specimen.
- 7.0 Place appropriate sex-wing length labels on insect pin (see Fig. 5).
- 8.0 Sort specimens within each species unit tray to sex-wing length for header-partial data entry.
- 9.0 Punch taxonomic names and numbers (categories 120, 130, 140, 150, 160), sex (category 320) and wing length (category 350) into paper tape.
- 10.0 Computer process data (steps 6.0 through 6.2.4 in Procedure I).
- 11.0 Determination of species names by comparison of specimens with type material (step 7.1 in Procedure I).
- 12.0 At this point the master file is basically finished (step 8.0 in Procedure I).

The result of Procedure II is rapid entry of specimen data by less repetitious typing of the same header data. By splitting and sorting the collection twice, typing of the locality header data and the taxonomic data is reduced to merely one entry and the chance of typing errors is minimized. With Procedure I, manual sorting takes place within each unit tray for header-partial data entry, and both sorting and typing are done on a small scale. With Procedure II, massive sorting takes place in one operation and less typing is needed, resulting in relatively fast entry of data.

PROCEDURE III (FIG. 7): NEW MATERIAL (SPECIMEN, SPECIMEN LABEL, AND
FIELD NOTE DATA) FROM THE STUDY'S CENTRAL AMERICAN EXPEDITIONS

- 1.0 Produce labels.
 - 1.1 Produce a bulk quantity of specimen number/SELGEM serial number labels (category 010) (step 1.0 in Procedure I).
 - 1.2 Produce specimen labels for the expedition.

FLOWCHART OF PROCEDURE II:
UNSORTED COLLECTIONS WHICH HAVE BEEN BORROWED FROM
PUBLIC (MUSEUMS) AND/OR PRIVATE SOURCES



*THIS STEP IS NOT ALWAYS MADE THROUGH
PUNCHED CARDS. SEE PROCEDURE I,
STEP 7.1. FOR EXPLANATION.

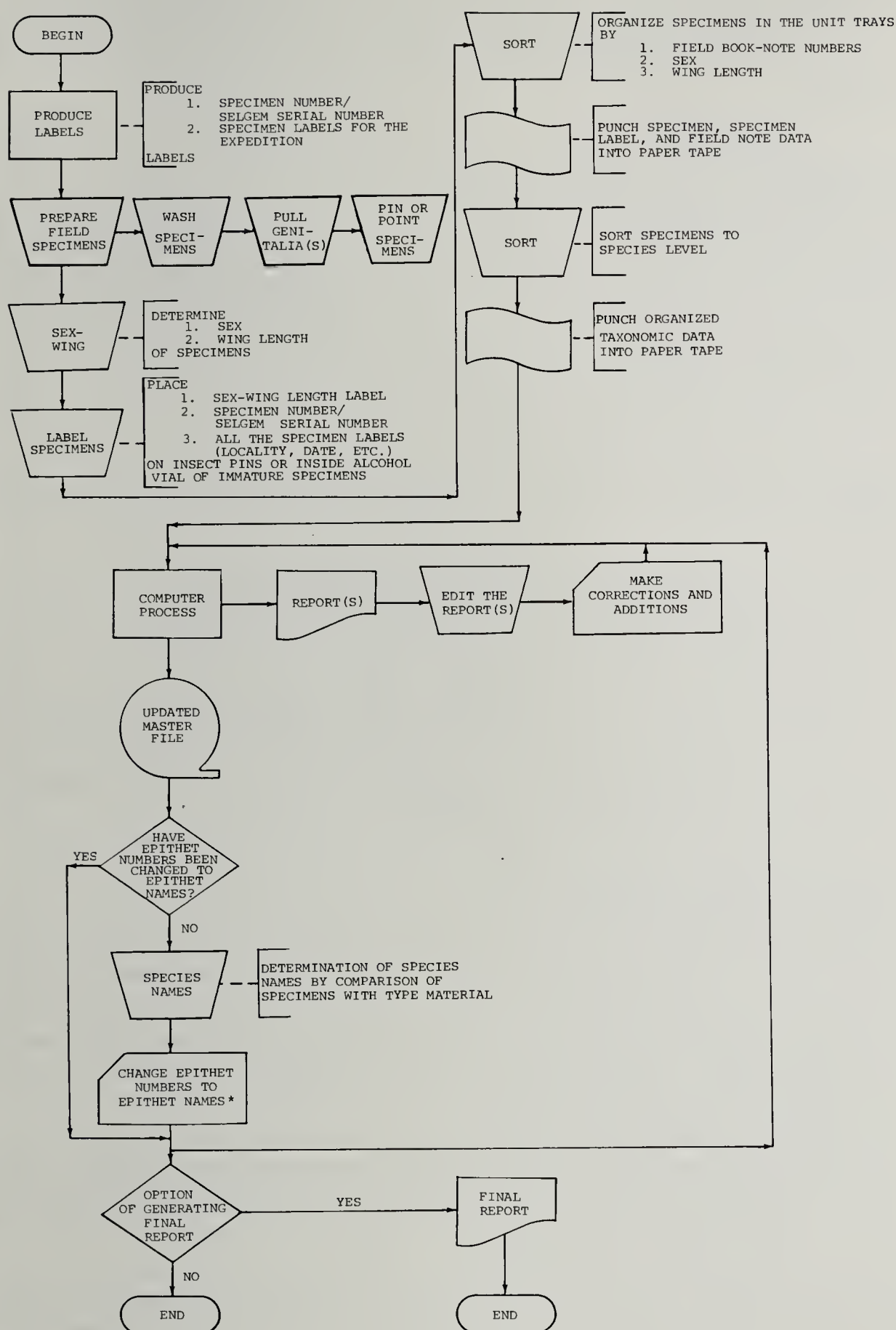
FIG. 6, Flowchart of Procedure II.

- 2.0 Prepare field specimens.
 - 2.1 Wash specimens in warm detergent water in a sonic cleaner; rinse specimens in warm or cool clear water under a faucet.
 - 2.2 When pinning or pointing specimens, pull the genitalia while the specimen is relaxed. Specimens are killed with ethyl acetate (see Lindroth, 1957, 1969).
- 3.0 Determine sex and wing length of specimens.
- 4.0 Label specimens.
 - 4.1 Place appropriate sex-wing length labels on insect pin (see Fig. 5).
 - 4.2 Place specimen number/SELGEM serial number (category 010) on insect pin or inside alcohol vial of immature specimens. (See Fig. 8 for the arrangement of the 2 labels on the pin. Note that at a glance you can see the sex, wing length, and specimen number. Data entry is faster if these 2 labels are placed at the top of the stack of labels.)
 - 4.3 Place all specimen labels on the insect pins or inside alcohol vials of immature specimens (i.e. locality with latitude-longitude, date, ecological label data, etc.).
- 5.0 Sort the specimens in the unit trays by field book-note numbers (category 820), sex (category 320) and wing length (category 350).
- 6.0 Punch label data and field note data into paper tape (i.e. locality, date, ecological label data, etc.).
 By having the specimens organized by field book-note numbers, sex, and wing length, all the label data will automatically be grouped for rapid header-partial data entry. The header record will be typed only once and will be followed by the partial records which will have the specimen number/SELGEM serial number and all the data for the specific individual specimen (i.e. ecological data, possibly specimen maturity [category 310] if the specimen is teneral).
- 7.0 Sort specimens to species level.
 - 7.1 Label each unit tray with the appropriate Genus name and epithet number.⁷
- 8.0 Punch taxonomic data (categories 120, 130, 140, 150, 160) into paper tape. By having the specimens organized taxonomically only one header record typing for each taxon is needed, followed simply by partial records containing the specimen number/SELGEM serial number.
- 9.0 Computer process data (steps 6.0 through 6.2.4 in Procedure I).
- 10.0 Determination of species names by comparison of specimens with type material (steps 7.1 in Procedure I).
- 11.0 At this point the master file is basically finished (step 8.0 in Procedure I).

In contrast to Procedures I and II where the latitude-longitude coordinates may not be (originally) in the file, Procedure III will already have had the coordinates incorporated into the file and therefore computer generated maps can be produced at any time.

This procedure is the most efficient of the 3 methods described, since the entire procedure from recording field notes to querying the file has been designed specifically for the Study with maximum efficiency and minimum usage of time.

FLOWCHART OF PROCEDURE III:
NEW MATERIAL (SPECIMEN, SPECIMEN LABEL, AND FIELD NOTE DATA)
FROM THE STUDY'S CENTRAL AMERICAN EXPEDITIONS



*THIS STEP IS NOT ALWAYS MADE THROUGH PUNCHED CARDS. SEE PROCEDURE I, STEP 7.1. FOR EXPLANATION

FIG. 7, Flowchart of Procedure III.

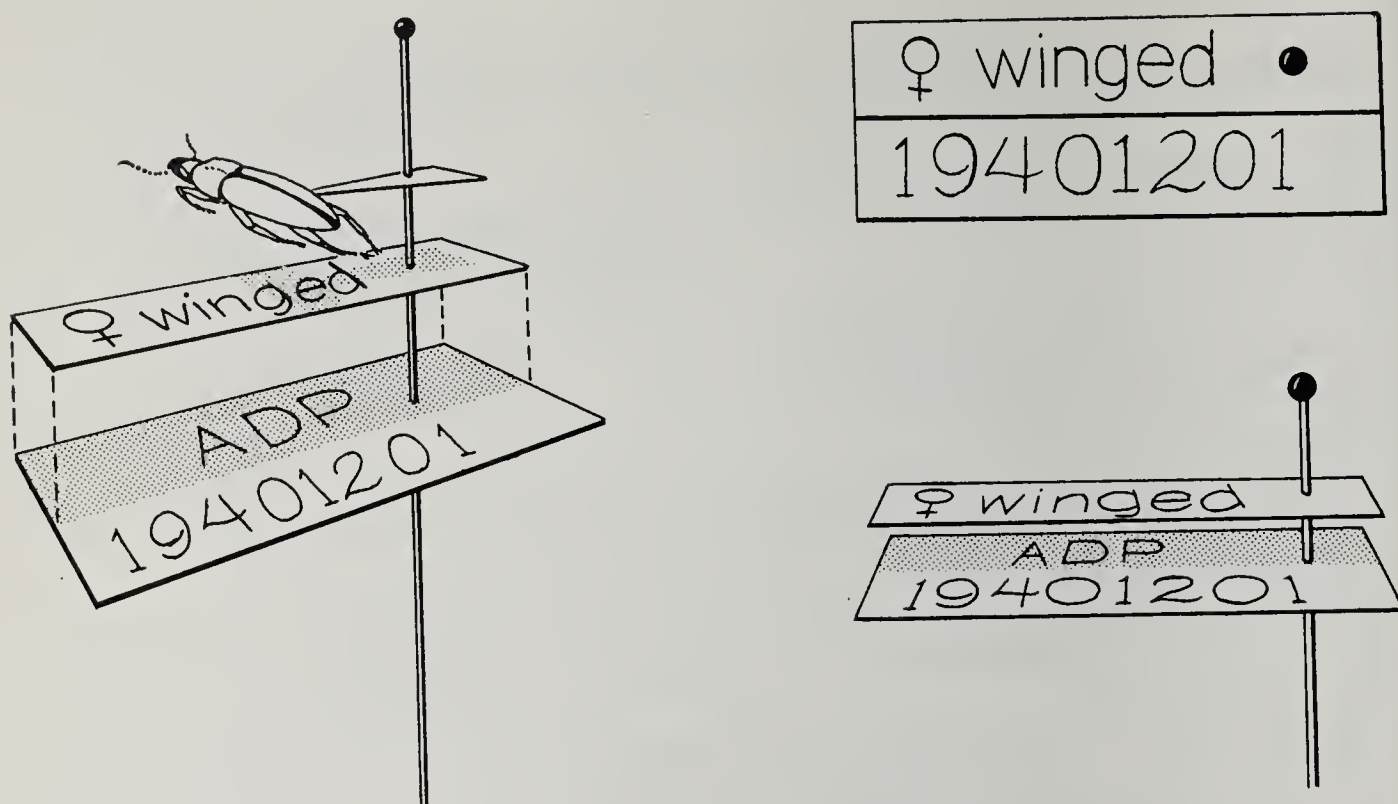


FIG. 8, Arrangement of Sex/Wing and ADP specimen number labels on insect pin.

DESCRIPTIONS OF CATEGORIES USED
AND HOW THEY ARE FORMATTED AND ENCODED
FOR FIELD NOTES AND FOR DATA PROCESSING⁸

For this Study most of the categories below are encoded into paper tape (except where noted) for subsequent transferral to magnetic tape. Other methods of data encodement are available, such as card punch machines and OCR (Optical Character Recognition).

Specimen Number (category 010)⁹

Every Central American Ground Beetle adult or immature specimen examined is randomly assigned a unique number not to exceed 8 digits. A label with this specimen number is attached to the insect pin for each adult or placed in the vial for each immature. Data can be traced back to each individual specimen, and/or data on any individual specimen can be retrieved, by using this unique number.

SELGEM identifies each stored record with an 8 character serial number. The *specimen number* and the *serial number* perform 2 completely separate functions; but for *this* Study the 2 numbers are designed to correspond to each other. Therefore, when reference is made to a specimen number, the specimen number will equal the serial number; but the serial number will *always* be an 8 digit (or character) number, e.g. 00001921, whereas the specimen number may be *smaller* than 8 digits, e.g. 1921. If the specimen number entry is less than 8 characters, a SELGEM program automatically generates leading zeros for the serial number. The specimen number labels are generated sequentially in intervals of one (e.g. 1, 2, 3, 4, etc.). To avoid confusion as to the meaning of the specimen label number, the acronym "ADP" was included on each specimen number label, as in Fig. 3. An improvement would be to add the name (acronym) of the depository, as in Fig. 4.

For other studies, lot numbers rather than individual specimen numbers may be used (such as all Ciidae captured in one bracket fungus, J. Lawrence, pers. comm.), if specimen uniqueness is not presently needed (e.g., sex or measurements of individual specimens). Subsequently, if individual specimen data are required, the lot number could be manipulated with computer techniques to produce individual specimen numbers.

Depository (category 020)

The acronym of the specimen depository is entered in this category. For the Study a 4 character acronym is assigned for a museum, unless a museum has an established acronym such as CAS, CNC, CSIRO, or MCZ. If specimens are borrowed from a private collection, a 6 character acronym is formed by combining the first 4 characters of the surname and the first letter of the first and middle names of the private collections' owner. If the surname has less than 4 characters, or if there is no middle initial, hyphens are used to replace the absent data characters (e.g. WHITDR; LEE-HH; KAVAD-).

Collection Transfer (category 030)

The historical continuum of the collections' depositories are entered in this category. Data in this category will allow future workers to easily re-assemble population samples which have been separated for deposit in diverse museums. The format used is: *ORIGINAL DEPOSITORY*, *semicolon*, *blank space*,¹⁰ and the *NEXT SEQUENTIAL DEPOSITORY* (e.g. DARLPJ;*MCZ;*USNM).

Specimen Loan-Return (category 040)

When a specimen in the Central American Study is loaned to an individual, the person's name, acronym of the museum with which the person is associated, date of loan, and date of return are entered in this category. The data for this category are entered into Hollerith 80 column IBM punch cards, not paper tape. Columns 1-15 are used for record identification. Data begin in column 16. The *NAME* of the individual to whom the loan is made is entered beginning in column 16. (See category 810 for the format used for entering person's names.) The *MUSEUM ACRONYM* to which the person is associated is entered in columns 50-53. The *DATE THE SPECIMEN WAS LOANED* is entered in columns 56-66. (See category 800 for the format used for entering dates.) The *DATE THE SPECIMEN IS RETURNED* is entered in columns 69-79. The format used from columns 50 to 80 is: UWIC**27*NOV*1973**02*DEC*1975*

Tribe (category 120)

Subtribe (category 130)

Genus (category 140)

Subgenus (category 150)

Species (category 160)

Subspecies (category 170)

The above 6 taxonomic levels are entered in separate categories. Individual levels can be easily queried, as well as combinations of the taxo-

nomic levels. Except for Species (category 160), all taxa are entered normally, i.e. BEMBIDIINI, for category 120, BEMBIDIINA, for category 130, etc. When the specimens are sorted to species, each species is assigned an epithet number, rather than an epithet name. This epithet number is entered in category 160, and remains in the computer file until comparison with types is completed. After the type comparisons are made, changing the epithet numbers to names is easily accomplished with common data processing techniques.

Identifier by Type Comparison (category 180)

A person's name and year of identification is entered in this category when the person is known to have identified the specimen on the basis of: 1) having seen the type; 2) having designated the lectotype or neotype; 3) having published on the taxon after seeing most types in the genus; or 4) having established reliable homotype material for comparative purposes. The format used is: *NAME* of the 180 Identifier (following the format explained under Collectors(s), category 810), *blank space*, abbreviation *DET* (determined), *blank space*, *YEAR* the identification was made, if available (e.g. DARLINGTON,P.J.*DET*1974).

Identifier by Keys or Unknown Method of Comparison (category 190)

If a determiner's name is included on the specimen label and it is not known whether the identifier used the methods described under category 180, then the identifier's name, year of identification, and determination are entered in this category. The format used is: *NAME* of the 190 Identifier (following the format explained under Collectors(s), category 810), *blank space*, abbreviation *DET* (determined), *blank space*, *YEAR* the identification was made, if available, *blank space*, and the *DETERMINATION* made (e.g. NEVERMANN,F.*DET *1931* CICINDELA*ARGENTATA*FABRICIUS).

Kind of Type (category 210)

If the specimen is labeled as a type, the kind of type is entered in this category, i.e. HOLOTYPE, PARATYPE, ALLOTYPE, SYNTYPE, HOMOTYPE, TOPOTYPE, NEOTYPE, LECTOTYPE, or PARALECTOTYPE.

Type Number (category 220)

Labels on type specimens frequently include the catalogue type number designated by the depository museum(s). This category includes the type number and is preceded by the depository's acronym. The format used is: *DEPOSITORY* where the type was catalogued, *blank space*, *TYPE NUMBER* (e.g. USNM*22101). If the specimen type labels indicate they have been catalogued differently in different depositories, then the depositories are listed sequentially, separated by a semicolon and a blank space (e.g. USNM*22101;*MCZ*94590).

Type Designator and Date (category 230)

The designator of the type followed by the year in which the determination was made, if available from the specimen label, is entered in this category. The format used is: *TYPE DESIGNATOR* (following the format explained under Collector(s) category 810), *blank space*, abbreviation *DET* (determined), *blank space*, *YEAR* the type was designated if available (e.g. LINDROTH,C.H.*DET*1974).

Original Type Name (category 240)

The name of the type (*exactly* as it appears on the specimen label) is entered in this category (e.g. AGRA DARLINGTONI STRANEO, A. DARLINGTONI, DARLINGTONI STRANEO, or simply DARLINGTONI).

Stage of Metamorphosis (category 300)

The stage of metamorphosis in which the specimen is preserved is entered in this category.^{11 12} Possible entries in this category are:

EGG
LARVAL*INSTAR,*FIRST
LARVAL*INSTAR,*SECOND
LARVAL*INSTAR,*THIRD
LARVAL*INSTAR,*FOURTH
LARVAL*INSTAR,*FIFTH
PUPA
ADULT

Specimen Maturity (category 310)

If the specimen is a fully hardened and pigmented adult, data are not entered. If the adult specimen is teneral, then one of the following 3 descriptive terms is entered in this category:

TENERAL-UNPIGMENTED
TENERAL-PARTIALLY*PIGMENTED
TENERAL-PIGMENTED

TENERAL-UNPIGMENTED is an extremely pale or nearly transparent and soft specimen preserved just after it emerged from the pupal stage; usually the abdomen is collapsed. TENERAL-PARTIALLY*PIGMENTED is a pale specimen which has not acquired its complete color; usually the abdomen is fully formed and not collapsed. TENERAL-PIGMENTED describes a specimen which has acquired full color, yet is still soft when collected.

Sex (category 320)

The sex of the specimen is entered in this category. Either FEMALE, MALE, two hyphens (--), or a question mark (?) is entered. The two hyphens are used if the sex is undeterminable prior to dissection. The question mark is used if the specimen is too badly damaged to determine the sex.

Specimen Length (category 330)

The "standardized body length" (Ball 1972) measured in millimeters (MM) is entered in this category. The format used is: a 3 *DIGIT NUMBER* (using leading zeros if necessary), a *decimal point*, one *DIGIT* to the right of the decimal, a *blank space*, abbreviation *MM* (e.g. 011.0*MM).

Specimen Width (category 340)

The specimen width measured in millimeters (MM) across the widest part of the elytra is entered in this category. The format used is: a 3 *DIGIT NUMBER* (using leading zeros if necessary), a *decimal point*, one *DIGIT* to the right of the decimal, a *blank space*, abbreviation *MM* (e.g. 006.5*MM).

Wing Length (category 350)

One of 3 descriptive terms (WINGED, SHORT-WINGED, or WINGLESS) is entered in this category to indicate the degree of development in the membraneous pair of metathoracic wings. WINGED refers to fully developed wings which extend beyond the stigma. SHORT-WINGED refers to a range of truncation at the stigma to a small veined pad. WINGLESS refers to a degree of development which ranges from a veinless pad to an apterous form.

Zoogeographic Region (category 400)

The zoogeographic regions as defined by Wallace (1876) and used by Darlington (1957) are entered in this category. For the Central American Study all entries are NEOTROPICAL, but as the Study evolves to include faunas other than Central America, the Study will expand to include the 6 other regions included in Wallace's Zoogeographic division (i.e. NEARCTIC, PALEARCTIC, ETHIOPEAN, ORIENTAL, AUSTRALIAN, and OCEANIA). The Zoogeographic Region will be used as an indicator in detecting distribution patterns.

Zoogeographic Subregion (category 410)

Category 410 is being reserved for Zoogeographic Subregion which will be used when the Study expands to include zoogeographic regions other than Neotropical.

Country (category 420)¹³

The Central American country in which the specimen was collected is entered in this category (i.e. BELIZE, COSTA RICA, EL SALVADOR, GUATEMALA, HONDURAS, NICARAGUA, PANAMA).

Department/District/Province (category 430)^{13 14}

The major administrative division of the country is entered in this category. The generic name of the administrative division (DEPARTMENT, DISTRICT, or PROVINCE) follows the specific administrative division (e.g. BOCAS DEL TORO PROVINCE).

General Locality (category 440)

A designated vicinity (a city, town, village, or river) that is at or near the actual field site and is easily located on major maps is entered in this category (e.g. PANAMA CITY; RIO ESCARREA).

Specific Locality (category 450)

The details which provide the closest proximity available for the collecting site are entered in this category. Notation's such as the number of miles or kilometers from the "general locality," or the name of a trail with the number of miles or kilometers from a given point are included. Miles will ultimately be converted to kilometers. For purposes of consistency in retrieval sorting, numbers are converted to a 3 digit number (using leading zeros), with one place to the right of the decimal point. Numbers are listed first, followed by the descriptive details (e.g. 030.8 MI N OF; or 000.3 KM SW OF BIOLOGICAL STATION ON FLORA TRAIL BETWEEN POSTS 11 and 12).

Latitude-Longitude (category 460)

The latitude-longitude coordinates of the specific or general locality are entered in this category. The format used for this Study is:

<i>Latitude</i>	<i>Longitude</i>
00*00*00*N.,	*000*00*00*E.
00*00*00*S.,	*000*00*00*W.

If the coordinates are not available, hyphens are inserted in place of the numbers. Leading zeros are used to complete the two and/or three digit numbers (e.g. 09*10*--N., *079*50*--W.).

Elevation (category 470)

Elevation in meters (M) is entered in this category. The format used is: a 4 *DIGIT NUMBER* (using leading zeros if necessary), *blank space*, and the code *M* (e.g. 0200*M). The format used for an elevation range is: a 4 *DIGIT NUMBER*, *blank space*, *TO*, *blank space*, 4 *DIGIT NUMBER*, *blank space*, and an *M* (e.g. 0300*TO*0450*M).

Holdridge Life Zone (category 500)

One Holdridge Life Zone name (Holdridge 1971) is entered in this category (e.g. TROPICAL DRY FOREST).

Community (category 520)

A broad expression describing the general surroundings of the microhabitat is entered in this category. Presently, the Study uses 18 community divisions, entered as either 4 or 5 character words: CAVE, CUL-L (cultivated land); FAL-F (fallow field); FORES (forest); GRASL (grassland); INHAB ("inhabitation"—in dwelling, parking lot, or similar disturbed area); LAKE; MARSH; PASTL (pastureland); POND; RIPAR¹⁵ (riparian);

ROADS (roadside, used also as trail); SAVAN (savannah); SCRUB; SEABE (seabeach); SEEPG (seepage); SUBPA (subparamo); and SWAMP.

Community Type (category 525)

When a specific kind of community descriptor is available, it is entered in this category. (e.g. If CUL-L was entered in category 520, then CORN, COFFEE, LETTUCE, etc. would be entered. If FORES was entered in category 520, then OAK, PINE, FICUS, etc. would be entered in this category.)

Community Type Narrative (category 526)

Narrative remarks concerning category 526 are entered in this category (e.g. CROPS HARVESTED, STALKS OLD AND DRY).

Size (category 530)

If the community category (520) is RIPAR or SEEPG, the width of the primary flood plain estimated in meters is entered in this category. The format used is: a 3 *DIGIT NUMBER* (using leading zeros if necessary), *blank space*, and an *M* (e.g. 020*M). The format used for a size range is: a 3 *DIGIT NUMBER*, *blank space*, *TO*, *blank space*, and an *M* (e.g. 020*TO*030*M).

Flow (category 535)

If the community category (520) is RIPAR or SEEPG, a descriptive term for flow is entered in this category. These descriptors are intended to create an image for the users. The Study uses the following at present:

- WATERFALL (steep gradient, water loses contact with substrate);
- CASCADING (steep gradient, water flow is extremely rapid, but does not lose contact with substrate, all "white water");
- RAPID (moderately steep, water moves swiftly, mix of "white water" and smooth surface, used only for large amounts of water);
- BABBLING (as in "rapid," but used for small amounts of water, such as rills and brooks);
- SLOW (low gradient, water obviously moves but has no "white water" on surface);
- STANDING (no gradient, water is not obviously moving, usually restricted to ponds in intermittent stream or river bed);
- DRY WATERCOURSE (any intermittent watercourse which is in its dry state).

Substrate (category 540)

If the community category (520) is LAKE, MARSH, POND, RIPAR, ROADS, SEABE, SEEPG, or SWAMP, a descriptive term for substrate is entered in this category. The terms used describing substrate are:

- BOULDER (large rock requiring an implement such as a crowbar to turn over);
- CLAY (firm, plastic, fine-grained earth, chiefly aluminum silicate);
- COBBLE (fist-size rounded stones which can easily be turned over with one hand);

GRAVEL (loose mixture of pebbles and rock fragments, coarser than sand, often mixed with clay, etc.);

HUMUS (brown or black product of the partial decay of leaves and other vegetable matter);

LATERITE (red, porous deposit containing large amounts of aluminum and ferric hydroxides, formed by the decomposition of certain rocks);

LEAF MOLD (rich soil consisting largely of decayed leaves);

LEAF LITTER (surface layer of the forest floor or elsewhere in which the leaves are slightly decomposed);

LOAM (rich soil composed of clay, sand, and some organic matter);

PEBBLE (small stone worn smooth and round, as by the action of water, about an inch or less in diameter);

SAND (loose, gritty particles of worn or disintegrated rock);

SILT (any earthy material composed of fine particles, as soil or sand, suspended in or deposited by water);

STONE (rock of relatively small size requiring two hands to turn over).

If there are two or more substrate terms, then on the field note form the descriptors are encircled with a 1, 2, 3, etc. to show the order of dominance. When encoding data the descriptors are listed in order of dominance and separated by a semicolon and a blank space (e.g. LEAF LITTER; *LEAF MOLD).

Substrate Narrative (category 541)

If the substrate category (540) requires narrative remarks, they are entered in this category (e.g. A THIN LAYER OF HUMUS ON ROCK SLABS).

Vegetation (category 545)

If the community category (520) is SAVAN, SCRUB, or SUBPA, then the dominant plant species is (are) entered in this category (e.g. HYPERICUM CARACASANUM).

If the community category (520) is LAKE, MARSH, POND, RIPAR, ROADS, SEABE, SEEPG, or SWAMP, the dominant associated plant species is (are) entered in this category (e.g. CAREX SPP.).

If there are 2 or more plant species, they are listed in order of dominance and separated by a semi-colon and a blank space (e.g. TYPHA LATIFOLIA; *CAREX VULPINOIDEA).

Structure (category 550)

If the community category (520) is ROADS, one of 3 terms describing the roadside structure (DITCH, MOUND, or FLAT) and narrative comments (if needed) are entered in this category. If narrative comments follow one of the terms, the terms are entered so that the narrative remarks begin on the 7th data character position (see glossary); the terms are followed by a

colon with either one or zero blank spaces following the colon, depending upon the number of characters in the term. The format possibilities are:

DITCH:narrative
MOUND:narrative
FLAT:*narrative

DITCH refers to drainage areas along roadways purposely dug for keeping the road bed dry; these ditches are usually maintained periodically for proper water drainage. MOUND refers to soil pushed up by the road graders and left along the road as a mound which is soon covered by pioneer vegetation. FLAT refers to that portion of the road bed which has been graded flat and left to pioneer plants.

Roadside Vegetation (category 555)

If the datum in the community category (520) is ROADS, one of 3 terms describing the zone for roadside vegetation (and narrative comments if needed) is (are) entered in this category. The 3 roadside vegetation zone terms are P-ZONE (pioneer zone), S-GROW (second growth), and CLIMAX (climax growth). If narrative comments follow one of the terms, the term is entered so that the narrative remarks begin on the 8th data character position; the terms are followed by a colon. The format possibilities are:

P-ZONE:narrative
S-GROW:narrative
CLIMAX:narrative

PIONEER ZONE refers to recently disturbed road margins with annual herbs and grasses and perhaps small shrubs and seedling trees; this zone is common on new and well maintained roadways. SECOND GROWTH refers to relatively undisturbed roadsides with small trees and shrubs and usually perennial herbs and grasses. CLIMAX GROWTH refers to mature stands of forest trees found normally only along old, poorly maintained roads or paths.

Bordered-by (category 560)

The community (and the community type if it applies) which *borders* the community (category 520) is (are) entered in this category. The bordered-by communities are entered as either a 4 or 5 character word (see category 520). If a community type (category 525) follows the bordered-by community, its entry begins in the 7th data character position. The format used is: *BORDERED-BY COMMUNITY* (first 4 or 5 characters), *colon, blank space* (only if the bordered-by community is a 4 character word), and the *COMMUNITY TYPE* (beginning in the 7th data character position). Examples of what may appear in category 560 if ROADS was entered in category 520 are: FORES:MIXED HARDWOODS, TREE FERNS or LAKE:*DYSTROPHIC.

Bordered-by (category 561)

If the community (category 520) is bordered by 2 different communities, the 2nd community and community type (if it applies) is (are) entered in this category using the same format entry as category 560.

Bordered-by (category 562)

If the community (category 520) is bordered by 3 different communities, the 3rd community and community type (if it applies) is (are) entered in this category using the same format entry as category 560.

Macrohabitat (category 580)

The gross community stratification and zonation, which is a coarse indicator of the horizontal and vertical layer in which ground beetles occur, is entered in this macrohabitat category. The 5 macrohabitat divisions presently used are CANOPY, UNDERCANOPY, SUPRA-EPIGEAN, EPIGEAN, and HYPOGEAN. CANOPY is entered if the specimen was associated with a tree crown in a forest community. UNDERCANOPY is entered if the specimen was found between the canopy and the ground in a forest community, such as associations with shrubs, vines, or fallen logs. SUPRA-EPIGEAN is associated with grass, shrubs, logs, and the like, in an open-area community such as cultivated land, fallow field, pasture-land, or savannah. EPIGEAN is entered for specimens found on the ground. HYPOGEAN is entered for underground associations.

Microhabitat (category 600)

A single word descriptor of the microenvironment with which the specimen is associated is entered in this category. The Study keeps a running glossary of microhabitat word descriptors so that consistency in recording data is maintained. Word descriptors commonly used in the Study are: ANT COLUMN, CARRION, CLAY, COBBLE, DEBRIS, DOMESTIC, DRIFT, DUNG, FLOWER, FOLIAGE, FOLIAGE-EPIPHYTE, FOLIAGE-FERN, FOLIAGE-GRASS, FOLIAGE-HERB, FOLIAGE-MOSS, FOLIAGE-SEDGE, FOLIAGE-SHRUB, FOLIAGE-TREE, FOLIAGE-VINE, FRUIT, FUNGI, GRAVEL, GUANO, HUMUS, LOAM, LEAF LITTER, MUD, NEST, ROOTS, SAND, SILT, SLASH, SOIL, STONE, SPIDER WEB, and WOOD.

Microhabitat Species Name (category 605)

The species name, or if not known, the common name of the microhabitat is entered in this category (if it applies). (e.g., If FOLIAGE is entered in category 600, and it is known that the species name is *Heliconia imbricata*, then HELICONIA IMBRICATA is entered in category 605; if STONE is entered in category 600, then the microhabitat species name category is left blank.)

Section of the Microhabitat (category 610)

The section of the microhabitat with which the specimen is associated is entered in this category. (For example, if WOOD is entered in category 600, possible entries would be TRUNK, BRANCH, TWIG. If FRUIT is entered in category 600, possible entries would be FLESH, SEEDS, SKIN. If FOLIAGE is entered in category 600, possible entries would be LEAF, PETIOLE, LEAF AXIL.)

Condition of Microhabitat (category 615)

Qualifiers for the condition of the microhabitat category are entered in this category (if applicable.) The Study keeps a running glossary of microhabitat word descriptors so that consistency in recording the data is maintained. (For example, if WOOD is entered in category 600, possible entries would be:

DRIFT
FALLEN, DEAD
FALLEN, LIVE
STANDING, DEAD
STANDING, LIVE

If FOLIAGE is entered in category 600, and LEAF is entered in category 610, possible entries would be:

WEBBED TOGETHER BY CATEPILLAR
UNROLLED YOUNG
DEAD

If FRUIT is entered in category 600, possible entries would be:

ADHERENT, DRY
ADHERENT, RIPE
ADHERENT, FERMENTING
DROPPED, DRY
DROPPED, UNRIPE
DROPPED, RIPE
DROPPED, FERMENTING)

Stratum of the Microhabitat Section (category 620)

The horizontal layer or section of the microhabitat is entered in this category. Common word entries are: ON, IN, UNDER, AMONG, FROM, UNDERSIDE (if FUNGI is entered in the microhabitat category), and BETWEEN (if LEAF AXIL is entered in the microhabitat section category). If WOOD is entered in category 600, possible entries would be:

BARK, ON
BARK, IN
BARK, UNDER
SAPWOOD
HEARTWOOD, IN

Activity of the Specimen (category 625)

The activity of the specimen prior to capture is entered in this category (e.g. FLYING, RUNNING, IN COPULA).

Microhabitat Remarks (category 630)

Microhabitat explanatory notes and remarks are entered in this category. If DEBRIS is entered in category 600, it is described in this category.

Collecting Method (category 700)

The method used to collect the specimen in its natural habitat is entered in this category. Collecting methods commonly used in the Study are BEATING, BERLESE FUNNEL, RAKING, SIFTING, SPLASHING, SWEEPING, and TREADING.

Attractants and Traps (category 720)

If specimens have been attracted or trapped, details of the method are entered in this category. This category differs from category 700 in that attractants and traps draw the insect away from its natural habitat. A single *word descriptor* of the attractant or trap is first entered in this category (e.g. BAIT, LIGHT, TRAP). A *colon* follows the attractant and trap word descriptor. If the word descriptor is less than 5 data characters, a *blank space* follows the colon (e.g. BAIT: *, LIGHT:). *Detailed word(s)* of the attractant and trap descriptor word is (are) entered in the 7th data character position; if there are more than one detail word, they are listed in order of dominance and separated by a blank space. (e.g.

LIGHT:UV,*TRAP
LIGHT:UV,*WHITE)

All labels which read "black light" are entered as "LIGHT:UV" Some attractant and trap data used in this Study are:

BAIT:*MALT, 2 DAYS OLD
BAIT:*MOLASSES
LIGHT:ARC-GLOBE
LIGHT:MERCURY VAPOR
LIGHT:UV
LIGHT:UV, WHITE
LIGHT:UV, TRAP
LIGHT:WHITE
TRAP:*CUT BRANCH
TRAP:*FRUIT FLY
TRAP:*FRUIT FLY, GLASS
TRAP:*PITFALL

Time of Activity (category 730)

The local time of day in which the specimen was observed active, in addition to the day phase is entered in this category. The *Diel* (24 hour) *Cycle* of recording time is used (using a leading zero, if necessary, to make a 4 digit number). A *blank space* follows the 4 digit time cycle and then either A, D, V, or N follows for Auroral, Diurnal, Vespereal, or Nocturnal (e.g. 1320*D).

Exposure (category 740)

The ILLUMINATION, GRADATION OF SLOPE, and the DIRECTION OF SLOPE are the 3 components of Exposure which are entered in this category. Each major category component has one data item entered from a choice of 3 or more.

They are:

ILLUMINATION
OPEN
PARTIALLY-OPEN
CLOSED
GRADATION OF SLOPE
FLAT
INTERMEDIATE
STEEP

They are formatted as:

ILLUM-
OPEN***
P-OPEN*
CLOSED*
GRAD-
FLAT**
INTER*
STEEP*

DIRECTION OF SLOPE	DIR-
NORTH	N
EAST	E
SOUTH	S
WEST	W
NORTHEAST	NE
NORTHWEST	NW
SOUTHEAST	SE
SOUTHWEST	SW

Illumination is entered in data character positions 1-6 as ILLUM-. Following the hyphen either OPEN***, P-OPEN*, or CLOSED* is entered in data character positions 7-13. "OPEN***" is entered if the microhabitat is openly exposed without any shade from trees or geological features, thus receiving the sun's full illumination, such as a tree trunk in the middle of cultivated land. "P-OPEN*" is entered if the microhabitat is in an open area receiving some source of shade during part of the day, such as a fallen log beneath a tree located in a fallow field. "CLOSED*" is entered if the microhabitat is situated in an area which does not allow penetration of sunlight, such as a log under a dense forest canopy.

Gradation is entered in data character positions 14-18 as GRAD-. Following the hyphen either FLAT**, INTER*, or STEEP* is entered in data character positions 19-24. "FLAT**" is entered for an exposed area that is at an approximate 10° level or less. "INTER*" is entered when an exposed area is approximately between an 11° and 22° slope. "STEEP*" is entered when an exposed area has a slope that is approximately 23° or more.

Direction is entered in data character positions 25-28 as "DIR-." Following the hyphen either N, S, E, W, NE, NW, SE, or SW is entered for the direction in which the slope faces the sun.

Example: ILLUM-OPEN***GRAD-INTER*DIR-SW

Topography (category 750)

The geological feature at which the specimen was collected is entered in this category. Some geological features used in the Study are CANYON, PLATEAU, RAVINE, and RIDGE. The geological feature is entered in the first 8 data character positions, followed by selected components and their descriptors. As an example, the selected components and descriptors for RAVINE are:

They are formatted as:

SECTION	SEC-
HEAD section of ravine	HEAD**
MID section of ravine	MID***
MOUTH section of ravine	MOUTH*
STRUCTURE	STR-
SHALLOW	SHALLOW*
SLOPING	SLOPING*
STEEP	STEEP***
SIZE of ravine	SIZE-
	000*M*
VEGETATION of ravine	VEG-narrative

"SEC-" is entered for ravine section in data character positions 9-12. Following the hyphen either HEAD**, MID***, or MOUTH* is entered in data character positions 13-18. The section of the ravine is subjectively determined by the field person.

"STR-" is entered for ravine structure in data character positions 19-22. Following the hyphen either SHALLOW*, SLOPING*, or STEEP*** is en-

tered in data character positions 23-30. These adjectives describe the walls of the ravine or canyon and are subjectively determined by the field person.

"SIZE-" is entered for the depth of the ravine in data character positions 31-35. A 3 *digit* estimated *number* is entered (using leading zeros if necessary) in data character positions 36-38. A *blank space*, an *M*, and another *blank space* follows in data character positions 39-41.

"VEG-" is entered in data character positions 42-45. The vegetation of the ravine follows in data character positions 46, 47, etc. (An example of how ravine and its components are entered is: RAVINE**SEC-HEAD**STR-SHALLOW*SIZE-005*M*VEG-BUNCH GRASS, FERN, PROSTRATE HERBS, FERNS ON WALL, GRASS AND HERBS ON FLOOR).

Date Captured (category 800)

The date on which the specimen was collected is entered in this category using a date format commonly used at the National Museum of Natural History for computerized data processing: The format used is: *DAY* (2 digits), *blank space*, *MONTH* (first 3 letters of the month), *blank space*, *YEAR* (4 digits). If any part of the date is not available, hyphens are used to replace the incomplete datum. If the value of the day of the month is 1 to 9, a leading zero precedes the one digit number (e.g. 10*JAN*1944; 01*DEC*1940; 20*AUG*----). A range of dates follows the same format, except for an addition of a *blank space*, *TO* and another *blank space* (e.g. 10*JAN*1974*TO*11*JAN*1974). The Central American Study uses this format not only for the data file but also for specimen labels, thus eliminating confusion and mistakes caused by the misinterpretation of the day, month, and/or year as explained in Stuckenberg and Irwin (1973) and Frank (1974).

Collector(s) (category 810)

The surname and initials of the specimen collector are entered in this category. The *surname* is entered first, followed by a *comma* and the *initials*. Each initial is followed by a period. Designations such as "Jr." or "Sr." follow the initials (e.g. DARLINGTON,P.J.JR.). Multiple entries of names are separated by a semi-colon and a blank space (e.g. DARLINGTON,P.J.JR.;*LINDROTH,C.H.). Titles, such as "Dr.," "Captain," are not entered.

Field Book-Note Numbers and Habitat Photograph (category 820)

The field *book* number, the field *note* number (which combined designate a particular microhabitat), and a code to indicate whether a photograph of the habitat was taken is entered in this category. Field notes are kept in a 6 1/2 by 9 inch field book binder which has its own 2 digit binder number (using a leading zero if necessary). Each page is called a "note" and contains data on one microhabitat and has its own 3 digit number (using a leading zero if necessary). The book and note numbers are combined into a hyphenated 6 character field book-field note number code. Following this code, a "P" preceded by a blank space is entered if a photograph was taken of the area in which the microhabitat occurs (e.g. 01-011*P).

When filling out the field note forms, the locality and constant data items are incorporated onto a lead form. The notes which follow contain

the specific microhabitat data. Lead notes and microhabitat notes have separate numbers. The last note in the series of microhabitat notes receives an encircled "E" (meaning end) located to the right of the 8 character code. The "E" is simply a double check on the numbers to make sure the variable microhabitat notes follow the correct lead or constant data items note.

Collector's Number (category 840)

Numbers on borrowed specimen labels which indicate that they are the collectors' field note number are entered in this category. These numbers are entered just as they appear on the label. If a large series of specimens from a collector is available, then the collector's numbers are formatted so that they are entered in specific data character positions (using leading zeros if necessary).

Expedition (category 850)

The name of the field expedition is entered in this category. Data relating to the expedition such as the museum or sponsor, the name of the expedition, and the date are included in this category (e.g. CMNH PANAMA EXPEDITION 1958).

Miscellaneous Remarks (category 999)

Explanatory notes and data which are not adaptable to any of the established categories are entered in this catch-all, miscellaneous remarks category (e.g. NO RAIN THE DAY SPECIMEN WAS COLLECTED; RIGHT HIND LEG MISSING WHEN COLLECTED; TEMPERATURE-054 F).

After a varied assortment of records have been accumulated from diversified sources (about 3,000 to 6,000 records), it is beneficial to review and analyze a listing of the miscellaneous remarks for patterns of consistent topics which require the creation of new standard categories.

ACKNOWLEDGEMENTS

I especially thank: Terry L. Erwin, my husband, who provided raw data, direction, and continued encouragement throughout the development of this paper, and David Bridge who generously shared his expert advice in matters relating to SELGEM, data processing, and biological problems. Also much appreciated is the cooperation, help, and advice extended to the Study by the *entire* NMNH-ADP Program staff during the tenures of James F. Mello (first year of the Study) and T. Gary Gautier. Hearty thanks are also extended to the following people who read early drafts of the manuscript and offered excellent ideas for its improvement: Robert T. Allen, David Bridge, Arthur A. Brigida, Terry L. Erwin, T. Gary Gautier, Patricia M. Juell, James F. Mello, and Donald R. Whitehead. Thanks are also extended to George Venable for preparing figures 3 and 7 and for graphic assistance and advice with the other figures. And, to the numerous people who have made helpful suggestions in the course of the Study, *thank you all*.

NOTES

¹SELGEM is the general purpose information management system developed at the Smithsonian Institution and used by the United States National Museum of Natural History and by many other museums. Information about application of SELGEM to museum collections is available from: Chief, The ADP Program, National Museum of Natural History, Smithsonian Institution, Washington, D. C. 20560. Technical information on SELGEM's operating system is available from: Manager, Information Retrieval and Indexing, Office of Computer Services, Arts and Industries, Smithsonian Institution, Washington, D. C. 20560.

²This manuscript is a 102 page draft entitled *FILE-BUILDING: A Manual for Data Banking via SELGEM*, labelled "Smithsonian Institution Procedures in Computer Sciences VOL. I, Num 2, FEBRUARY 1973." Copies are available through Information Retrieval and Indexing, address in footnote 1.

³See section of paper entitled "Procedures for Preparing and Gathering Source Data."

⁴See the section of this paper entitled "Descriptions of categories used and how they are formatted and encoded for field note and for data processing."

⁵Note that the categories are grouped by subject matter and that intervals were designed for inclusion of subject related categories.

⁶The ADP specimen number/SELGEM serial number label (Fig. 3) which is on each Central American Ground Beetle specimen used in this Study is produced as follows: 1) A computer program was written to generate consecutive numbers below the acronym "ADP" (Automatic Data Processing). (The printing is done on white computer paper. Care is taken to use a *good* ribbon so that a sharp contrast of black ink on white paper is achieved); 2) The printout which has 4 characters per centimeter is photo reduced to make a negative with 9 characters per centimeter; 3) The negative is then used to print copies on 100% rag content paper; and 4) The labels are cut and applied to each pinned specimen or put into the alcohol specimen vial before data are entered into the computer.

⁷To maintain a continuous work flow for the Study's needs, each species is assigned a generic name and epithet number to be used until the specimens can be compared to type material, when epithet numbers are changed to names.

⁸The field note form and the category formats should also be useful to non-computer oriented Coleopterists as an example of how field notes can be structured to maximize efficiency and accuracy in gathering and recording field data.

⁹The specimen number serves the same purpose as do catalogue numbers assigned to type specimens in large insect collections or to catalogue numbers assigned to specimens in disciplines other than Entomology.

¹⁰Here and elsewhere in the text, an asterisk (*) is used to denote a blank space; on the field note form (Fig. 1) the blank space is indicated by a Greek *delta*.

¹¹The first phase of the Study will deal mostly with adults. More advanced phases of the Study will allow for in depth natural history studies involving all life history stages, at which time this category may need to be expanded to allow for greater details of each stage.

¹²Some of the following examples are in columnar form, as visual support for stressing comparative formats for data character positions or to visually illustrate involved formatting. For those examples which do not need visual support, horizontal lists are used.

¹³The placement of such areas as the CANAL ZONE are entered arbitrarily according to personal preference, and may not accurately reflect political attachment. The Central American Ground Beetle Study arbitrarily enters PANAMA in the Country category, and the CANAL ZONE in the Department/District/Province category.

¹⁴Frequently the Department/District/Province datum is not available on the specimen labels and has to be incorporated into the file subsequent to the original data entry. A special report can be produced from the computer, listing all the locality data in any requested order. In a report made for the Central American Ground Beetle Study, the Country, Department/District/Province, General Locality, and Specific Locality were listed in columns in alphabetical order. If the province had been entered into the file, it was listed in the province column of the report; if not, a blank area occurred between the country and the general locality. The province for each general locality need be looked up only one time. Then by using data processing methods, province is entered only once followed by all the specimen numbers/SELGEM serial numbers pertaining to specimens from this general locality. The computer then inserts in the file the appropriate province for each specimen record.

¹⁵For the Central American Ground Beetle Study, *riparian* specifically refers to a relationship with a river, stream, creek, brook, or run.

REFERENCES CITED

- ANON. 1962. Automatic data processing glossary. Bureau of the Budget (USA). 62 p.
- ANON. 1970. Vocabulary for information processing, X3.12-1970. Federal information processing standard 11. American National Standards Institute, Inc. 128 p.
- ANON. 1972. Federal information processing standards publication (FIPS Pub 20). National Bureau Standards (USA). 11 p.
- BALL, G. E. 1972. Classification of the species of the *Harpalus* subgenus *Glanodes* Casey (Carabidae: Coleoptera). Coleopt. Bull. 26(4):179-204.
- CHENHALL, R. G. 1975. Museum cataloging in the computer age. The American Association for State and Local History, Tennessee, 261 p.
- CREIGHTON, R., AND J. J. CROCKETT. 1971. SELGEM: A system for collection management. Smithsonian Institution Information Systems Innovations, 2(3):1-24.
- CREIGHTON, R., P. PACKARD, AND H. LINN. 1972. SELGEM RETRIEVAL: A general description. Smithsonian Institution Procedures in Computer Sciences 1(1):1-38.
- CREIGHTON, R., AND P. PACKARD. 1974. Computer assisted information management: Getting oriented. Smithsonian Institution Procedures in Computer Sciences 1(2):1-19.
- CREIGHTON, R., AND E. ROBINSON. 1975. Computer assisted information management: Input and file maintenance. Smithsonian Institution Procedures in Computer Sciences 1(4):i-v, 1-36.
- DARLINGTON, P. J., JR. 1957. Zoogeography: The geographical distribution of animals. John Wiley and Sons, Inc., New York, 675 p.
- FRANK, J. H. 1974. The Fourth of July, or what's in a date? A plea. Coleopt. Bull. 28(1):48, 50.
- GAUTIER, T. G. 1974. A report on the use of SELGEM at the National Museum of Natural History. Association of Systematics Collections Newsletter 2(1):12-17.
- HOLDRIDGE, L. R., W. C. GRENKE, W. H. HATHEWAY, T. LIANG, AND J. A. TOSI, JR. 1971. Forest environments in tropical life zones, a pilot study. Pergamon Press Ltd., New York, 747 p.
- LINDROTH, C. H. 1957. The best method for killing and preserving beetles. Coleopt. Bull. 11:95-96.
- LINDROTH, C. H. 1969. The ground-Beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 1. Opusc. Ent. Suppl. 35:1-48.
- MANNING, R. B. 1969. Automation in museum collections. Proc. Biol. Soc. Wash. 82:671-686.
- SHETLER, S. G. 1974. Demythologizing biological data banking. Taxon 23(1):71-100.
- STUCKENBERG, B. R., AND M. E. IRWIN. 1973. Standards for entomological labels. Bull. Ent. Soc. Amer. 19:164-168.
- SQUIRES, D. F. 1966. Data processing and museum collections: A problem for the present. Curator 9(3):216-227.
- SQUIRES, D. F. 1970. An information storage and retrieval system for biological and geological data. Curator 13(1):43-62.
- WALLACE, A. R. 1876. The geographical distribution of animals. Macmillan, London, Vol. 1:503 p.

GLOSSARY

The following terms and definitions are either from glossaries listed in the references cited or have been altered or developed for clarity in reading this paper.

Category (sensu SELGEM)

One or more data elements. In the SELGEM system a category is identified by a 3 digit number.

Data

A representation of facts or concepts suitable for communication, interpretation, or processing by human or automatic means. Singular form, DATUM.

Data Bank (or Data Base)

A comprehensive collection of *files* of data. For example, several *categories* comprise a *record*, a set of records forms a *file*, and the files used by a researcher or group of mutual interest researchers are known as their *data bank*.

Data Character Position

When a data item(s) is (are) formatted, the field of data must start and end in specified locations. Each location within the field has a special identifying address known as the data character position.

Data Element

A basic unit of identifiable and definable information. It has an identifying name and value or values for expressing a specific fact (e.g. Country, Date, Collecting Method, Community, Microhabitat).

Data File

A collection of related data records organized in a specific manner and treated as a unit.

Data Item

The expression of a particular fact in a category, e.g. "Neotropical" may be the data item in the Zoogeographic Region category.

Field

A specified area within a category used for a particular data item.

Formatted Information

An arrangement of information into discrete units and structures, to facilitate access and processing; in contrast, narrative information is arranged according to rules of grammar.

Header-Partial Data Entry

A method used to speed the data input process by reducing keystrokes. All constant data are typed into one "header" record, and the variable data are entered into "partial" records which are strung behind the "Header" record (which for this Study is on paper tape). A SELGEM program adds the header data to each partial record producing complete specimen records in the file.

Leading Zeros

When a field for a data item is formatted for a specific number of numeric positions (digits), leading zeros are added to the left of the number when the number has less than the required number of digits. For example, if a field requires 4 digits, the number 10 would have 2 leading zeros added, thus it would read: 0010.

Master File

A collection of related records which is altered infrequently and supplies basic data for processing operations.

Optical Character Recognition (OCR)

A method of encoding data for computer processing. Data are typed on a special typewriter which has an optical character recognition element.

The OCR typed copy is machine read through the use of light-sensitive devices.

Record (Unit Record, *sensu* SELGEM)

A collection of categories of data, treated as a unit. For the Central American Ground Beetle Faunal Study, a specimen record contains all the categories of data for a particular specimen.

SELGEM

The acronym for *SELf-GEnerating Master*. SELGEM is the information management system developed at the Smithsonian, consisting of over 25 general purpose computer programs. This storage, management, and retrieval system was designed especially for collection management in museums and to handle natural history data.

Serial Number

In the SELGEM system, an 8 character number used to identify a record.

Source Data

For this paper, "source data" refers to data from specimens, specimen labels, and field notes.

Unit Tray

For the non-entomologist who may read this paper, a unit tray is a small topless box with a special bottom that firmly holds the insect pins, and in which the insects are housed. Several unit trays of insects are housed in an insect drawer. Several drawers are housed in an insect cabinet.

